DOCUMENT RESUME

ED 395 981 TM 025 140

AUTHOR Buras, Avery
TITLE Descriptive versus Predict ve Discriminant Analysis: A Comparison and Contrast of the Two Techniques.
PUB DATE 26 Jan 96
NOTE 32p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, January 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Behavioral Science Research; Comparative Analysis; *Discriminant Analysis; *Effect Size; *Group Membership; *Research Methodology; Social Science Research
IDENTIFIERS *Descriptive Discriminant Analysis; *Predictive Discriminant Analysis

ABSTRACT

The use of multivariate statistics in the social and behavioral sciences is becoming more and more widespread. One multivariate technique that is commonly used is discriminant function analysis. This paper compares and contrasts the two purposes of discriminant analysis, prediction and description. Using a heuristic data set, a conceptual explanation of both techniques is provided with emphasis on which aspects of the computer printouts are essential for the interpretation of each type of discriminant analysis. Initially, discriminant analysis was designed to predict group membership, given a number of continuous variables. It also is used to study and explain group separation or group differences. Descriptive discriminant analysis has been used traditionally as a followup to a multivariate analysis of variance. The explanation of the differences in these two approaches includes discussion of how to: (1) detect violations in the assumptions of discriminant analysis; (2) evaluate the importance of the omnibus null hypothesis; (3) calculate the effect size; (4) distinguish between the structure matrix and canonical discriminant function coefficient matrix; (5) evaluate which groups differ; and (6) understand the importance of hit rates in predictive discriminant analysis. An appendix presents a syntax file from the Statistical Package for the Social Sciences. (Contains 7 tables and 20 references.) (SLD)

Descriptive Verses Predictive Discriminant Analysis:

A Comparison and Contrast of the Two Techniques.

Avery Buras

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA. January 26, 1996.

ABSTRACT

The use of multivariate statistics in the social and behavioral sciences is becoming more and more widespread. One multivariate technique that is commonly used is discriminant function analysis. The present paper will compare and contrasts the two purposes of discriminant analysis, prediction and description. Using a heuristic data set, a conceptual explanation of both techniques is provided with emphasis on which aspects of the computer printouts are essential for the interpretation of each type of discriminant analysis.

To honor a reality in which we believe that any given effect can have one or many causes and in which any given cause could have one or multiple effects, it is vital for the researcher to understand the application of multivariate statistics (Thompson, 1986). Dolenz (1993) reported that even though this is becoming more widely accepted in research, many graduate programs in the social sciences carry statistics courses that focus on univariate analysis and culminate only with a detailed look at analysis of variance (ANOVA). Empirical studies of present practice also indicate that univariate analysis, and particularly ANOVAs, are still the predominant statistical method that is chosen in the behavioral sciences (Elmore & Woehlke, 1988; Goodwin & Goodwin, 1985.)

Studies in the social sciences comparing two or more groups very often measure subjects on several dependent variables (Stevens, 1993). Statistical techniques which examine two or more dependent variables simultaneously are referred to as multivariate. For example, a researcher may want to investigate the impact of four teaching techniques (Methods A, B, C, and D) upon the four subjects (dependent variables) of reading comprehension, arithmetic, spelling and problem solving. After randomly assigning the students to one of the four classes, each subject area is measured using an intervally scaled instrument.

A graduate student who has just finished a course in ANOVA, may be tempted to analyze the above data by doing four one way ANOVAs, one ANOVA for each dependent variable. If statistical significance is noted, this student would then do post hoc tests for each statistically significant ANOVA. Fish (1986) noted two reasons why this is undesirable. First, doing four different ANOVAs inflates the possibility of a Type I "experimentwise" error. Thompson (1994) reports that most researchers are familiar with "testwise alpha" or the probability of making a Type I error for a given hypothesis. However, little attention is given to the probability

of making a Type I error anywhere in the study, i.e., the "experimentwise" error rate. The "experimentwise" error for four one way ANOVAs is conceptually about 4 times the testwise alpha level ($\alpha_{TW} = .05$) or approximately 20% for perfectly uncorrelated dependent variables.

If the dependent variables in the above example are in fact perfectly uncorrelated the "Bonferroni inequality" would be the more precise way of calculating the "experimentwise" error. Applying the "Bonferroni inequality" to perfectly uncorrelated variables, the chances of making a Type I error (α_{TW}=.05) somewhere in our experiment would be approximately 18.55% (Thompson, 1994).

$$\begin{aligned} \alpha_{EW} &= 1 - (1 - \alpha_{TW})^K \\ &= 1 - (1 - .05)^4 \\ &= 1 - (.95)^4 \\ &= 1 - (.8145) \\ \alpha_{EW} &= .1855 \end{aligned}$$

Researchers can control this "experimentwise" error by using the "Bonferroni correction" (Thompson, 1994). The "Bonferroni correction" involves the calculation of a new testwise alpha level, computed by dividing the testwise alpha by the number of hypotheses. However, this lowered alpha level could lead to less statistical power or Type II error. Fish (1988) and Maxwell (1992) have both provided data sets which illustrate the paradoxical effect of failing to identify statistically significant results when univariate tests are used inappropriately when multivariate tests should have been employed.

Thompson (1994) noted that "the use of the 'Bonferroni correction' does not address the second (and more important) reason why multivariate methods are so often vital, and so even with this correction univariate methods usually still remain unsatisfactory" (p. 12). This "more important reason" that Thompson (1994) refers to is the second reason reported by Fish (1988), i.e., the use of several univariate tests does not have the ability to reflect the reality which we

believe exists. However, multivariate methods have the ability to reflect the reality of the data from which the researcher is working. Just as independent variables can interact to produce statistically significant results, so too can dependent variables interact to produce statistically significant results (Thompson, 1994). This interaction of dependent variables can be detected by the use of multivariate techniques. The use of multivariate techniques can take into account the intercorrrelations of the independent and dependent variables. Whatever the case, multivariate statistics can take into consideration these interactions and intercorrelations (Thompson, 1994).

In the present paper, the multivariate technique that will be focused upon is discriminant function analysis. Specifically, the paper will compare and contrast descriptive discriminant function analysis (DDA)and predictive discriminant function analysis (PDA). A data set will be used to explain and illustrate the similarities and differences of these two techniques. While the data used in the paper are real data from another research project, the research question has been changed in this paper for ease of explanation. This fictional research questions used to illustrate DDA and PDA was referred to above. Does teaching method A, B, C, or D affect performance in reading comprehension, arithmetic, spelling and/or problem solving?

Overview

Initially, discriminant analysis was designed to predict group membership, given a number of continuous variables (Dolenz, 1993). For example, if incumbent candidates were running for office and wanted to predict whether or not they were going to be re-elected, they could gather information on previous incumbent candidates and whether or not they were elected. To predict their re-election the candidate may choose variables such as the condition of the economy, number of foreign crises, tax rates, and any other variables that may be important to predict re-election. From a previous sample of senators, a linear discriminant function (LDF)

can be derived such that a new individual can be placed into one of the categories of re-elected or not re-elected (Huberty, 1975), and any senator could predict his or her own individual chances.

The second purpose of discriminant analysis is to study and explain group separation or group differences (Huberty & Wisenbaker, 1992). The use of DDA techniques to describe group differences began to be used in the 1960's (Huberty, 1975). Traditionally, DDA techniques have been used as a follow-up to a multivariate analysis of variance (MANOVA) (Huberty & Morris, 1989). In DDA, a set of weights are obtained and a linear combination of a set of response variables is computed to maximize between-group separation while minimizing within-group variance (Klecka, 1980). This minimization of within-group variance and the maximization of between-group variance by the use of a set of weights is also employed in ANOVA, Multiple Regression and t-Tests (Thompson, 1991).

Discriminant analysis basically consists of a set of intervally scaled variables and a set of grouping or categorical variables. To determine which set of variables is the predictor variables and which set is the criterion variables, the research question is required. Each research situation determines the direction of causation and thus whether or not PDA or DDA is to be used (Klecka, 1980). If group membership is being used to predict or explain scores on the continuous variables, DDA is used. If the scores on the continuous variables are used to predict group membership, PDA is used. In a DDA the group variables are treated as independent variables while the dependent variables are the continuous variables. In the example given above, the independent variables are the teaching techniques while the dependent variables are the scores in the four subject areas. If we were trying to predict which students respond better to each of the four teaching techniques we could use the scores on the four tests to predict class

membership. In this PDA, the dependent variables are group membership and the independent variables are the interval scores on the four tests.

Assumptions of DDA and PDA

Klecka (1980) described seven mathematical assumptions of discriminant analysis. In order for a discriminant analysis to be conducted, the following seven assumptions must be met:

1) two or more groups which are mutually exclusive;

2) at least two subjects per group;

3) any number of discriminating (continuous) variables can be used provided that the number of cases exceeds the number of variables by more than two;

4) discriminating variables are measured at the interval level;

5) no discriminating variable may be a linear combination of other discriminating variables;

6) the covariance matrices for each group must be (approximately) equal, unless other special formulas are used;

7) each group has been drawn form a population with a multivariate normal distribution on the discriminating variables.

Interpretation of DDA Results

When interpreting the results of a DDA three questions drive our analysis of the results. First, do the groups differ? Second, which groups differ? Third, if they do differ, on which dependent variables do they differ? Historically, a MANOVA would be run and if statistically significant results were found, a DDA would be run as a post hoc test. The primary run of a one-way MANOVA program prior to a DISCRIMINANT program is unnecessary, however, given that a one-way MANOVA and discriminant analysis are the same thing (Huberty & Wisenbaker,

1992). In fact, the SPSS MANOVA and DISCRIMINANT c nmands yield essentially the same information on the computer printouts (Dolenz, 1993). Interested readers are encouraged to "prove" this for themselves by running the SPSS syntax file presented in Appendix 1. In discriminant analysis, statistics reported which are of interest and will be discussed in the present paper include canonical correlations, eigenvalues, and Wilks lambda, as well as standardized coefficients, structure coefficients, and an evaluation of group centroids (Dolenz, 1993).

Before looking at the results, and addressing the three questions, it is first important to consider whether the basic assumptions of discriminant analysis have been met. Using a DISCRIMINANT program, it is possible to test the assumptions associated with discriminant analysis (Huberty & Barton, 1989). Univariate homogeneity of variance is tested in SPSS using Cochran's test of homogeneity of variance and Bartlett-Box F. The results for our data suggest that there is no statistically significant difference in the variances of the dependent variables across the four teaching techniques.

Insert Table 1 About Here

Stevens (1992) reports that except for rare examples, multivariate normality can be detected by methods assessing for univariate normality. However, caution is advised; since univariate normality is a necessary but not a sufficient condition for multivariate normality we cannot conclude definitively that we have multivariate normality even if we do have univariate and bivariate normality. However, if there was a statistically significant and noteworthy difference in the univariate normality, we could not proceed any further.

The second assumption that is tested is the homogeneity of the variance/covariance matrices for each dependent variable across the four groups. SPSS uses Boxes M as the test for

homogeneity of the variance/covariance matrices. Included in Table 2, which has been taken directly from the computer print-out, are the variance/covariance matrices for each group and the pooled variance/covariance matrix as well as an F test for homogeneity of variance/covariance. Since the F statistic was not statistically significant and the test is very powerful, we can conclude that the assumption that the matrices be approximately equal has been met (Klecka, 1980). Since there was no statistically significant difference in the variance/covariance matrices for our data, we can proceed to answering our three questions.

Insert Table 2 About Here

Our first question can be answered by inspecting the omnibus null hypothesis or the multivariate test of statistical significance. The omnibus null for our data refers to the question, do the different teaching techniques produce differences on the variables of arithmetic, reading comprehension, spelling and/or problem solving? For our data, Wilks' multivariate test of significance will be used, although there are three other methods are also used to calculate statistical significance for a MANOVA (Heausler, 1987). One-way MANOVA and DISCRIMINANT results across the different teaching techniques indicated a statistically significant difference in our data [F=2.346 (12,455.36), p=.006] as shown on Table 3. The computer printout also reports univariate F-ratios for the four dependent variables.

Insert Table 3 About Here

The second and third questions to be answered refer to which groups differ and on which dependent variables do they differ. We can answer these question by examining the discriminant functions. Before proceeding with these questions, it is important to understand what a discriminant function is and how many discriminant functions are possible. Discriminant

function scores are a linear combination of the discriminating variables (intervally scaled variables) which are formed to satisfy certain conditions: the discriminant function is the set of weights applied to the response variables to compute these discriminant function scores. The first condition is that the discriminant functions are derived in order to maximize the separation of the groups (between-group variance) while minimizing the dispersion of scores within each group (within-group variance) (Hubertv, 1984).

The number of discriminating functions derived in discriminant analysis is based on the number of groups and the number of discriminating variables. The number of functions equals the number of groups minus one or the number of discriminating variables, whichever is smaller (Huberty, 1975). The coefficients that compose the first function are derived to maximize the differences between the groups. The coefficients for the second function are also derived to maximize the dispersion of the groups with the added condition that the values on the second function are not correlated with values on the first function (Klecka, 1980). The third function is derived in a way which maximizes group differences without being correlated with the first or second functions. This process continues up to the number of unique functions which can possibly be derived, with some of the latter functions being trivial and lacking statistical significance (Dolenz, 1993).

Since statistical significance is largely an artifact of sample size (Cohen, 1994), other means of evaluating whether or not a researcher has found meaningful results have been suggested. Effect size has been suggested as an alternative to statistical significance or to be used along with statistical significance (Cohen, 1994). One effect size statistic derived from discriminate function analysis is the canonical correlation coefficient, a measure of association between the groups and the discriminant function (Klecka, 1980). By squaring the canonical

correlation coefficient, a statistic analogous to eta^2 is derived. In the example presented above, the first canonical correlation is .3706, making eta^2 equal to .1373 or 13.73%. The researcher could then conclude that a noteworthy amount of variance in scores on the discriminating variables is predictable form group membership.

Insert Table 4 about here

The most common test for statistical significance is based on Wilks' lambda (Klecka, 1980). Wilks' lambda is also an "inverse" measure, analogous to $1\text{-}eta^2$, with a maximum of one and a minimum of zero. An effect size for a DDA can be calculated by subtracting the value of Wilks lambda from 1. In tables 3 and 4 above, Wilks' lambda is reported as .85325. Therefore, effect size could also be calculated by 1 - .85325 making the effect size equal to .14675 or 14.675%.

Another statistic that is reported in discriminant analysis and can be seen in Table 4 is an eigenvalue. Although eigenvalues cannot be interpreted directly, the relative magnitude of the eigenvalues can be used to describe the relative value of each function (Klecka, 1980). The function with the largest eigenvalue is the largest discriminator, and the functions with the smaller eigenvalues are the least powerful at discriminating the groups. In Table 4, Function 1 has an eigenvalue of .159 and Function 2 has an eigenvalue of .011. From these two eigenvalues, we can conclude Function 1 discriminates 14 times better than Function 2.

Now that we have concluded that there is a statistically significant and meaningful difference in our four teaching methods, and that these differences lie only in Function 1, we need to turn our attention to the question, which groups differ? By looking at Table 5, and examining the canonical discriminant functions evaluated at the group centroids, we can see the

group 1, 2, and 3 are approximately at the same points on Function 1 and group 4 is a considerable distance from groups 1, 2 and 3. We can therefore conclude that group 4 members are effected most by that teaching method.

Insert Table 5 about here

Now that we know that Function 1 discriminates group 4 from groups 1, 2, and 3, we need to ascertain what variables compose function one. This is done by examining the standardized canonical discriminant function coefficients and the structure matrix of each function. The standardized coefficient gives that variable's relative unique contribution to calculating the discriminant score Klecka, 1980). Since standardized coefficients are conceptually analogous to beta weights in regression, they cannot be interpreted alone. Standardized coefficients are derived with the relative contribution of all variables being considered simultaneously (Thompson, 1992). Dolenz (1993) writes,

> A problem with standardized coefficients arises when variables have high intercorrelations, causing the intercorrelating variables to "compete" for weighted values. Conceptually, a variable that would carry a high weight if considered alone may be "blocked" by a variable sharing the same discriminating information. Interpretation of this blocked variable's standardized coefficient would cause the erroneous conclusion that it was not an important contributing variable. (pp. 11-12)

While standardized coefficients consider all variable contributions to the function simultaneously, structure coefficients are bivariate correlations and therefore, are not affected by relationships with other variables (Klecka, 1980). Structure coefficients explain which variables

combine to compose the function. Structure coefficients can range from -1.0 to +1.0 since they are simple correlations. By noting those variables which make up the largest portion of the function, we can attempt to name the function (Klecka, 1980).

Insert Table 6 about here

By examining the structure matrix in Table 6 we can see that READING COMP correlates .88 with Function 1 and SPELLING correlates .68 with the Function 1. It is the responsibility of the researchers to rely on their own creativity and their knowledge of the literature to name and describe each function. Since Function 1 is composed mainly of reading comprehension and spelling, it could be concluded that teaching method D influences score in reading comprehension and spelling, i.e., "verbal" areas.

Interpretation of Results of PDA

As stated earlier, the original purpose of discriminant analysis was the prediction of group membership (Huberty & Wisenbaker, 1992). The focus in this analysis changes from the description of the influences of group membership on the scores on intervally-scaled variables to a focus on group classification accuracy or the percentage of cases correctly classified based on using intervally-scaled scores as predictor variables. How then do we decide which group a case actually belongs in? Huberty (1994) noted that the "decision or classification or assignment rule that is commonly used is based on the *maximum likelihood principle*: Assign a unit to the population in which its observation vector has the greatest likelihood of occurrence" (p. 45).

In discriminant analysis it is possible to graph the function scores for each individual subject onto a P dimensional space, where P refers to the number of functions that are calculated Klecka, 1980). Since one of the conditions placed upon function scores is to maximize between

group variance while minimizing within group variance, each group's members will tend to cluster about the group centroid. Conceptually, a subject is classified based upon their position in the P dimensional space, with assignment going to the group whose centroid is the closest distance from that particular subject's discriminant score vector.

When a subject is classified into the closest group based upon this distance, this assignment is also implicitly based upon assigning it to the group for which it has the highest probability of belonging (Klecka, 1980). One probability that can be calculated is a "typicality probability" (Huberty, 1994). SPSS DISCRIMINANT produces a "typicality probability" table denoted by P(D.G), which refers to the probability of having the discriminant score vector given membership in the stated group. Klecka (1980) describes a "typicality probability" as the chance that a case that far from the group centroid could actually belong to that group. A small typicality probability implies a greater distance of the discriminant score vector from the stated group centroid (Huberty & Wisenbaker, 1992). For example, in Table 7 case 191 has a 31.10% chance of coming from it's stated group membership of group 3. Case 3 on the other hand, has a 97.10% chance of coming from it's stated group, 4. Huberty and Wisenbaker (1992) note that an object associated with a small typicality probability of less than .10 could be considered a possible outlier. They also suggest possible ways to deal with potential outliers.

Insert Table 7 about here

Another type of probability that is calculated is a "posterior probability," denoted by P(G D), which refers to the probability of belonging to any group, given a particular score vector (Huberty, 1994). Each subject is given a set of "posterior probabilities," one posterior probability for each group. By definition these sets of "posterior probabilities" must sum to 1.00 (Huberty,

1994; Klecka, 1980). The reason the "posterior probabilities" sum to 1.00 across groups for each subject can be illustrated with the following extreme case. It is possible that any subject could have 100% chance of belonging to group 1. This would mean that by definition this subject would have a 0% chance of belonging to groups 2, 3, or 4. A subject is assigned to the group which has the highest probability of belonging. Again, classification on the largest of these values also is equivalent to using the smallest distance (Klecka, 1980). "Posterior probabilities" can be calculated for each group, but SPSS reports only the two highest values for each subject.

It is often clear which group a case should be assigned to based upon the typicality probabilities or posterior probabilities. For example, it is clear based upon the posterior probabilities that case number 191 "belongs" in group 4. However, it may not be readily apparent which group some cases belong. For example, cases 1, 2, 3, 190 and 192 all have relatively similar close "posterior probabilities." The data used in our study could be considered to have a low level of discrimination, therefore, group membership may not be "neatly" concluded. When this is the case, the subjects are likely to have similar probabilities for each group. Klecka (1980), encourages researchers to be cautious about decisions surrounding these types of cases, especially when there is evidence that the assumption of multivariate normality has not been met.

The number of cases correctly predicted by the classification functions is called the hit rate, the total focus of PDA. The higher the hit rate, the better the functions predict group membership. Also included in Table 7 are the classification results. In this particular study, roughly 44.13% of the cases were correctly classified based up the functions derived from our sample. While it would be desirable to have a higher hit rate, with our classification functions we can predict better than chance (25%) what the group membership was. An example of a poor

hit rate would be a PDA with only two groups and a classification result of 50%. By chance alone, we would have a 50% probability of predicting group membership correctly. Therefore, a hit rate of 50% using predictive information results in no improvement over prediction using no information.

The classification functions that were derived in the present paper were based upon an equal probability of being assigned in a particular group. If we had prior knowledge that a particular group had 70% of the cases, and the remaining three groups had 10% each, we would want the evidence to be strong that a member assigned to the smaller groups actually belonged there. This can be accomplished by adjusting the posterior probabilities by taking into account these prior probabilities (Klecka, 1980).

Another instance in which the prior probabilities should be taken into consideration is when the study involves relatively high stakes. Klecka (1980) refers to this as the cost of misclassification. His example pertains to the determination of whether a patient has malignant or benign cancer. The cost of misclassifying a person with a malignant cancer into the benign cancer group is readily apparent. The researcher would want the evidence to be overwhelming that cases actually belong to the benign group before they are classified. This added confidence in the classification can be accomplished by adjusting for prior probabilities (Klecka, 1980).

Internal vs. External Hit Rates

A shortcoming of our present data is that the typicality probabilities printed by the SPSS DISCRIMINANT program are based on an "internal analysis" (Huberty & Wisenbaker, 1992). This method, the most common method used in the behavioral science, uses the data to formulate a classification function and then classifies the same data with the obtained rule (Huberty, Wisenbaker, & Smith, 1987). This so-called "apparent hit rate," typically yields

classifications results better than a "true hit rate". A "true hit" rate refers to the classification of a future sample based upon an empirically derived rule or function. The reasoning behind the positive bias of an "apparent hit rate" is analogous to the maximization of R in regression. Since the weights are obtained by optimizing the variance of the sample at hand, sampling error idiosyncrasies in the data will influence positively the internal hit rate (Huberty, 1994).

Another method for identifying the "true hit rate" would be an "external analysis" such as a "holdout method" or a "leave-one-out method" (Huberty & Wisenbaker, 1992). One way of carrying out an external classification is to randomly split the available data into two smaller samples. With one of the sub-samples, calculate a classification function and then use the discriminant functions to predict the membership of the other sub-sample. Typically, one sub-sample is larger and the larger sub-sample is used to derive the classification function. The "true hit rate" is determined by classifying the sub-sample that has been left out. Huberty, Wisenbaker and Smith (1987) have called this external classification method the "holdout method," since part of the sample has been held out.

Another method of calculating an external classification function is called the "leave-one- out method" (L-O-O) (Huberty, Wisenbaker & Smith, 1987). This method involves deleting one subject and determining a linear classification function based upon the remaining N-1 subjects. These linear classification functions are used to classify the deleted unit into one of the groups. This process is carried out N amount of times (Huberty, Wisenbaker & Smith, 1987).

There are limitations to these alternate ways of calculating hit rates. For further information on the draw-backs and benefits of calculating these two types of external hit rates the reader is directed to Huberty, Wisenbaker and Smith (1987). The detailed presentation of

these methods of hit rate calculation are beyond the scope of this work, and not because these methods are not important.

A Final and Important Distinction Between DDA and PDA

Generally, the adding of variables to a statistical analysis does not take away from effect size, and often increases uncorrected effect sizes. This is also true for DDA (Huberty, 1994). However, in PDA, fewer variables can yield greater classification accuracy, whereas in DDA, fewer variables cannot yield greater discrimination (Huberty, 1994). Thompson (1995) stresses that this is an important point and that this apparent paradox emphasizes the importance in distinguishing DDA from PDA.

One option that is available on statistical packages such as SPSS is the plotting of territorial maps (Thompson, 1995). These plots indicate the boundaries of the groups and include notations as to the location of each subject in the variable space. Some subjects may be close to the group centroids of the groups on these territorial maps, while other subjects may be "fence-riders" or lie just within the boundaries of a particular territory. The paradoxical effect happens because the subjects, in the data set with more variables, will always move on the average closer to their respective group centroids, which results in a decreased Wilks' lambda (increasing the effect size). However, some subjects could move only slightly further from their group centroid into a wrong group. For example, when a variable is added, a given subject who was originally a correctly-classified "fence rider" could move considerably closer to its respective group centroid while three other subjects who were initially correctly classified but also "fence riders" could move a very small distance into the wrong group upon the addition of new predictor variables. The net result is an increase in effect size but the undesirable effect of a decrease in the hit rate (Thompson, 1995).

Conclusion

Discriminant analysis techniques are being widely used in educational research (Huberty & Barton, 1989). The present paper was not intended to be an exhaustive survey of discriminant analysis, but rather, has attempted to familiarize the reader with the important information that may be encountered when trying to read and understand research articles that have used a discriminant analysis. Emphasis was also placed on the reading and understanding of computer generated printouts.

It is hoped that the reader at this point has an understanding of the differences between (PDA) and (DDA). Also, the reader has been encouraged to understand how to detect violations in the assumptions of discriminant analysis, how to evaluate the importance of the omnibus null hypothesize, how to calculate the effect size, how to distinguish between the structure matrix and canonical discriminant function coefficient matrix, how to evaluate which groups differ, and the importance of hit rates in predictive discriminant analysis.

References

Cohen, J. (1994). The earth is round (p < .05). American Psychologist, 49(12), 997-1003.

Dolenz, B. (1993, January). Descriptive discriminant analysis: An application. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 355 274)

Elmore, P.B., & Woehlke, P.L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Research, 17(9), 19-20.

Fish, L.J. (1988). Why multivariate methods are usually vital. Measurement and Evaluation in Counseling and Development, 21, 130-137.

Goodwin, L.D. & Goodwin, W.L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14(2), 5-11.

Heausler, N.L. (1987). A Primer on MANOVA Omnibus and Post Hoc Tests. Paper presented tat the annual meeting of the Southwest Educational Research Association, Dallas, TX. (ERIC Document Reproduction Service No. ED 281 852).

Huberty, C. J. (1975). Discriminant analysis. Review of Educational Research, 45(4), 543-598.

Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. Psychological Bulletin, 95(1), 156-171.

Huberty, C.J. (1994). Applied Discriminant Analysis. New York: Wiley and Sons.

Huberty, C.J. & Barton, R.M. (1989). An introduction to discriminant analysis. Measurement and Evaluation in Counseling and Development, 22, 158-168.

Huberty, C.J. & Morris, J.D. (1989). Multivariate analysis versus multiple univariate analysis. Psychological Bulletin, 105, 302-308.

Huberty, C. & Wisenbaker, J. (1992). Discriminant analysis: Potential improvements in typical practice. In B. Thompson (Ed.), Advances In Social Science Methodology (Vol. 2, pp. 169-208). Greenwich, CT: JAI Press.

Klecka, W.R. (1980). Discriminant Analysis. Beverly Hills, CA: Sage.

Maxwell, S. (1992). Recent developments in MANOVA applications. In B. Thompson (Ed.), Advances in Social Sciences Methodology (Vol. 2, pp. 137-168). Greenwich, CT: JAI Press.

Stevens, J. (1993). Applied Multivariate Statistics for the Social Sciences. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Thompson, B. (1986, November). Two reasons why multivariate methods are usually vital. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis.

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in Counseling and Development, 24(2), 80-95.

Thompson, B. (1992, April). Interpreting regression results: beta weights and structure coefficients are both important. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 344 897).

Thompson, B. (1994, February). Why multivariate methods are usually vital in research: Some basic concepts. Paper presented at biennial meeting of the southwestern Society for Research in Human Development, Austin.

Thompson, B. (1995). Review of Applied discriminant analysis by C.J. Huberty. Educational and Psychological Measurement, 55, 340-350.

Table 1

SPSS Printout: Univariate Homogeneity of Variance Tests

Variable ..MATH	
Cochrans C(44,4) =	.34039, P = .123 (approx.)
Bartlett-Box F(3,37496) =	1.83860, P = .138
Variable ..SPELLING	
Cochrans C(44,4) =	.28700, P = .828 (approx.)
Bartlett-Box F(3,37496) =	.21474, P = .886
Variable ..READING COMP	
Cochrans C(44,4) =	.27727, P = 1.000 (approx.)
Bartlett-Box F(3,37496) =	.27527, P = .843
Variable ..PROBLEM SOLVING	
Cochrans C(44,4) =	.28839, P = .797 (approx.)
Bartlett-Box F(3,37496) =	.64890, P = .584

Table 2

SPSS Printout: Variance/Covariance Matrix for Each Group and Statistical Significance Test for Homogeneity of Variance/Covariance Matrices

Cell Number .. 1
Variance-Covariance matrix

	MATH	SPELLING	READING COMP	PROBLEM SOLV
MATH	89.736			
SPELLING	-5.471	118.943		
READING COMP	-9.107	-60.786	74.593	
PROBLEM SOLV	-55.476	-8.914	28.667	107.168

Determinant of Covariance matrix of dependent variables = 29003630.49329
LOG(Determinant) = 17.18293

- -

Cell Number .. 2
Variance-Covariance matrix

	MATH	SPELLING	READING COMP	PROBLEM SOLV
MATH	52.720			
SPELLING	4.760	125.378		
READING COMP	-7.200	-39.969	58.685	
PROBLEM SOLV	-50.560	-3.000	5.400	97.680

Determinant of Covariance matrix of dependent variables = 14675514.96235
LOG(Determinant) = 16.50169

- -

Cell Number .. 3
Variance-Covariance matrix

	MATH	SPELLING	READING COMP	PROBLEM SOLV
MATH	72.349			
SPELLING	-4.635	153.606		
READING COMP	19.794	-59.822	75.171	
PROBLEM SOLV	-58.454	1.025	-23.117	92.330

Determinant of Covariance matrix of dependent variables = 22943989.27570
LOG(Determinant) = 16.94857

- -

Table 2 Continued

Cell Number . 4
Variance-Covariance matrix

	MATH	SPELLING	READING COMP	PROBLEM SOLV
MATH	48.819			
SPELLING	-3.956	137.283		
READING COMP	-3.005	-56.417	62.661	
PROBLEM SOLV	-25.265	5.995	15.389	74.436

Determinant of Covariance matrix of dependent variables = 14386655.81828
LOG(Determinant) = 16.48181

- -

Pooled within-cells Variance-Covariance matrix:

	MATH	SPELLING	READING COMP	PROBLEM SOLV
MATH	62.266			
SPELLING T	-3.150	135.179		
READING COMP	-.265	-55.622	66.981	
PROBLEM SOLV	-41.559	00.734	8.916	87.882

Determinant of pooled Covariance matrix of dependent vars. = 21708953.83377
LOG(Determinant) = 16.89324

- -

Multivariate test for Homogeneity of Dispersion matrices

Box's M = 30.62654
F WITH (30,35928) DF = .96934, P = .513 (Approx.)
Chi-Square with 30 DF = 29.10579, P = .512 (Approx.)

- -

Table 3

SPSS Printout: Multivariate Test of Statistical Significance (Omnibus Null) and Univariate Tests of Statistical Significance

Analysis of Variance -- design 1

EFFECT .. TEACHING METHOD
Multivariate Tests of Significance (S = 3, M = 0, N = 85)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.14825	2.26145	12.00	522.00	.009
Hotellings	17023	2.42100	12.00	512.00	005
**Wilks	.85325	2.34585	12.00	455.36	.006**
Roys	.13732				

- -

EFFECT .. TEACHING METHOD (Cont.)
Univariate F-tests with (3,175) D. F.

Variable	Hypoth SS	Error SS	Hypoth MS	Error MS	F	Sig of F
MATH	319.908	10896.528	106 635	62 266	1.71259	166
SPELLING	1725.308	23656 301	575.102	135 179	4.25438	006
READING COMP	1455 55	11721.751	485.185	66 981	7.24358	000
PROBLEM SOLV	215.65	15379 333	71 883	87 882	81795	486

Table 4

SPSS Printout. Canonical Discriminant Functions

Fcn	Eigenvalue	Pct of Variance	Cum Pct	Canonical Corr	After Fcn	Wilks' Lambda	Chi-square	df	Sig
					0	.853250	27.614	12	.0063
1*	.1592	93.51	93.51	.3706	1	.989071	1.912	6	.9276
2*	.0107	6.28	99.79	.1028	2	.999637	.063	2	.9689
3*	.0004	.21	100.00	.0190					

* Marks the 3 canonical discriminant functions remaining in the analysis

Table 5

SPSS Printout: Canonical discriminant functions evaluated at group means (group centroids):

Group	Func 1	Func 2	Func 3
A	-.34550	00484	-.03372
B	-.38590	-.20331	.01855
C	-.35163	.14844	.01947
D	.43371	-.00287	00038

Table 6

SPSS Printout: Standardized Canonical Discriminant Function Coefficients and Structure Matrix:

Standardized canonical discriminant function coefficients:

	Func 1	Func 2	Func 3
MATH	.23501	1.18908	-.00699
SPELLING	20192	.04661	.26590
READING COMP	.79409	-.22643	.49135
PROBLEM SOLV	-.25202	.75391	.85607

Structure matrix:

Pooled within-groups correlations between discriminating variables
and canonical discriminant functions
(Variables ordered by size of correlation within function)

	Func 1	Func 2	Func 3
READING COMP	.88187 *	-.17094	.43544
SPELLING	.67587 *	.14323	-.01530
MATH	.38027	.76486 *	-.49908
PROBLEM SOLV	-.29313	05987	.91889 *

* denotes largest absolute correlation between each variable and any discriminant function.

Table 7

SPSS Printout: Typicality Probability and Hit Rates (Classification Results):

Case Number Group	Actual Group	Highest Probability	P(D/G)	P(G/D)	2nd Highest Group	P(G/D)	Discrim Scores
1	UNGRPD	3	.9390	.3019	1	.2893	-.7569
							.4791
							-.3443
2	3 **	1	.9884	.2550	3	.2492	.0050
							-.0579
							-.0361
3	4	4	.9710	.2536	3	.2504	.0469
							-.0169
							.2994
....							
190	4 **	2	.9412	.2665	4	.2628	.1201
							-.5716
							-.0414
191	3 **	4	.3110	.7619	3	.1650	1.1578
							-.3814
							-.2750
192	1	1	.1577	.3235	2	.3134	-1.7310
							-.1416
							-1.8392

Classification results -

Actual Group	No. of Cases	Predicted Group Membership 1	2	3	4
Group 1	36	3	10	13	10
		8.3%	27.8%	36.1%	27.8%
Group 2	26	2	12	6	6
		7 7%	46.2%	23.1%	23.1%
Group 3	36	6	7	12	11
		16.7%	19.4%	33.3%	30.6%
Group 4	81	1	13	15	52
		1.2%	16.0%	18.5%	64.2%
Ungrouped cases	13	2	3	1	7
		15.4%	23.1%	7.7%	53.8%

Percent of "grouped" cases correctly classified: 44.13%.

Appendix 1

SPSS Syntax File For MANOVA and DISCRIMINANT programs.

```
MANOVA
 math probsolv readcomp spelling BY method(1 4)
 /DISCRIM RAW STAN ESTIM CORR ROTATE(VARIMAX) ALPHA(1)
 /PRINT SIGNIF(MULT UNIV EIGN ) SIGNIF(EFSIZE) CELLINFO(CORR)
CELLINFO(COV)
 HOMOGENEITY(BARTLETT COCHRAN BOXM)
 /NOPRINT PARAM(ESTIM)
 /METHOD=UNIQUE
 /ERROR WITHIN+RESIDUAL
 /DESIGN.
DISCRIMINANT
 /GROUPS=method(1 4)
 /VARIABLES=math probsolv readcomp spelling
 /ANALYSIS ALL
 /PRIORS EQUAL
 /STATISTICS=MEAN STDDEV UNIVF BOXM COEFF RAW CORR COV GCOV TCOV
TABLE
 /PLOT=CASES
 /CLASSIFY=NONMISSING POOLED.
```